Evaluation of measurements by the method of least squares

Lars Nielsen

Danish Institute of Fundamental Metrology (DFM), Lyngby, DK.[1]
LN@dfm.dtu.dk

Abstract

In this paper, a general technique for evaluation of measurements by the method of Least Squares is presented. The input to the method consist of estimates and associated uncertainties of the values of measured quantities together with specified constraints between the measured quantities and any additional quantities for which no information about their values are known a priori. The output of the method consist of estimates of both groups of quantities that satisfy the imposed constraints and the uncertainties of these estimates. Techniques for testing the consistency between the estimates obtained by measurement and the imposed constraints are presented. It is shown that linear regression is just a special case of the method. It is also demonstrated that the procedure for evaluation of measurement uncertainty that is currently agreed within the metrology community can be considered as another special case in which no redundant information is available. The practical applicability of the method is demonstrated by two examples.

1 Introduction

In 1787, the French mathematician and physicist Laplace (1749–1827) used the method of Least Squares to estimate 8 unknown orbital parameters from 75 discrepant observations of the position of Jupiter and Saturn taken over the period 1582–1745. Since then, the method of Least Squares has been used extensively in data analysis. Like Laplace, most people use a special case of the method, known as unweighted linear regression. The calculation of the average and the standard deviation of a repeated set of observations is the most simple example of that. The unweighted regression analysis is based on the assumptions that the observations are independent and have the same (unknown) variance. In addition, the linear regression is based on the assumption that the observations can be modelled by a function that is linear in the unknown quantities to be determined by the regression analysis. For most measurements carried out in practice, none of these assumptions can be justified. In order to evaluate the result of a general measurement, in which some redundant information has been obtained, one therefore has to apply the method of Least Squares in its general form.

This paper describes how measurements can be evaluated by the method of Least Squares in general. The paper is based on an earlier work of the author [2] but includes

[1]Address: Building 307, Matematiktorvet, DK-2800 Lyngby, Denmark

several new features not published before as well as practical examples from the daily work at DFM. An alternative approach is described in [6].

2 Measurement model

In a general measurement, a number $m > 0$ of quantities is either measured directly using measuring instruments or known a priori, for example from tables of physical constants etc. The (exact) values of these m quantities are denoted $\boldsymbol{\zeta}$

$$\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_m)^T .$$

Due to measurement uncertainty, the values $\mathbf{z}$ obtained by the measurement (or from tables etc.)

$$\mathbf{z} = (z_1, ..., z_m)^T$$

are only estimates of the values $\boldsymbol{\zeta}$. The standard uncertainties of the estimates z_i,

$$u(z_i) \quad , \quad i = 1, \ldots, m,$$

are determined in accordance with the GUM [1] and depend on the accuracy of the instruments and the reliability of any tabulated value used. In general, some of the estimates z_i may be correlated. If $r(z_i, z_j)$ is the correlation coefficient between the estimates z_i and z_j then the covariance $u(z_i, z_j)$ between these two estimates is given by

$$u(z_i, z_j) = u(z_i) r(z_i, z_j) u(z_j).$$

Because of the uncertainty, the estimates $\mathbf{z}$ can be considered as an outcome of a m-dimensional random variable $\mathbf{Z}$ with expectation $\boldsymbol{\zeta}$ (the exact values of the quantities) and covariance matrix $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} = u(\mathbf{z}, \mathbf{z}^T) = \begin{pmatrix} u^2(z_1) & u(z_1, z_2) & \cdots & u(z_1, z_m) \\ u(z_2, z_1) & u^2(z_2) & \cdots & u(z_2, z_m) \\ \vdots & \vdots & \ddots & \vdots \\ u(z_m, z_1) & u(z_m, z_2) & \cdots & u^2(z_m) \end{pmatrix} .$$

In addition to the m quantities for which prior information is available either from direct measurement or from other sources, a general measurement may involve a number $k \geq 0$ of quantities for which no prior information is available. The values of these quantities are denoted by

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T .$$

In general, the values $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ are constrained by a number n of physical or empirical laws. These constraints may be written in terms of an n-dimensional function

$$\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta}) = \begin{pmatrix} f_1(\boldsymbol{\beta}, \boldsymbol{\zeta}) \\ f_2(\boldsymbol{\beta}, \boldsymbol{\zeta}) \\ \vdots \\ f_n(\boldsymbol{\beta}, \boldsymbol{\zeta}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad , \quad k \leq n < m + k. \tag{2.1}$$

It is assumed that $f_i : \Omega \to R$, $i = 1 \ldots n$, are differentiable functions (with con-

tinuous derivatives) defined in a region $\Omega \subset R^{k+m}$ around $(\boldsymbol{\beta}, \boldsymbol{\zeta})$. As indicated in (2.1), the number n has to be larger than or equal to the number k of quantities for which no prior information is available; otherwise some of the values $\boldsymbol{\beta}$ cannot be determined. In addition, the number n of constraints has to be smaller than the total number $k+m$ of quantities involved; otherwise the values of $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ would be uniquely determined by the constraints and no measurements would be needed.

The estimates $\mathbf{z}$, the covariance matrix $\boldsymbol{\Sigma}$ and the n-dimensional function $\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta})$ are the input to the general Least Squares method. It should be stressed that no probability distribution has to be assigned to the input estimates $\mathbf{z}$. On the contrary, if a probability distribution has been assigned to an estimate, it should be used to calculate the mean value and the variance of the estimate which should then serve as input to the Least Squares method.

Like any other covariance matrix, the covariance matrix $u(\mathbf{z}, \mathbf{z}^T) = \boldsymbol{\Sigma}$ is positive semi-definite. Otherwise, at least one linear combination $\mathbf{x}^T\mathbf{z}$ of the estimates $\mathbf{z}$ would have *negative* variance $u(\mathbf{x}^T\mathbf{z}, \mathbf{z}^T\mathbf{x}) = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}$. In the following it is assumed that $\boldsymbol{\Sigma}$ is positive definite and therefore non-singular.

3 Normal equations

Least Squares estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\zeta}}$ of the values $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ are found by minimizing the chi-square function

$$\chi^2(\boldsymbol{\zeta}; \mathbf{z}) = (\mathbf{z} - \boldsymbol{\zeta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\zeta})$$

subject to the constraints

$$\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta}) = 0.$$

It is convenient to solve this minimization problem by using Lagrange multipliers [5]: If a solution $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})$ to the minimization problem exists, the solution satisfies the equation

$$\nabla_{(\boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\lambda})} \Phi(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}, \boldsymbol{\lambda}; \mathbf{z}) = \mathbf{0}$$

where

$$\Phi(\boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\lambda}; \mathbf{z}) = (\mathbf{z} - \boldsymbol{\zeta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\zeta}) + 2\boldsymbol{\lambda}^T \mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta})$$

for a particular set of Lagrange multipliers $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^T$. By taking the gradient of the function Φ, the following $n + m + k$ equations in $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}, \boldsymbol{\lambda})$ evolve:

$$\begin{aligned} \nabla_{\boldsymbol{\beta}} \mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})^T \boldsymbol{\lambda} &= \mathbf{0}, \\ -\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \hat{\boldsymbol{\zeta}}) + \nabla_{\boldsymbol{\zeta}} \mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})^T \boldsymbol{\lambda} &= \mathbf{0}, \\ \mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}) &= \mathbf{0}, \end{aligned} \tag{3.1}$$

where

$$\nabla_{\boldsymbol{\beta}} \mathbf{f} = \begin{pmatrix} \frac{\partial f_1}{\partial \beta_1} & \cdots & \frac{\partial f_1}{\partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial \beta_1} & \cdots & \frac{\partial f_n}{\partial \beta_k} \end{pmatrix} \quad \text{and} \quad \nabla_{\boldsymbol{\zeta}} \mathbf{f} = \begin{pmatrix} \frac{\partial f_1}{\partial \zeta_1} & \cdots & \frac{\partial f_1}{\partial \zeta_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial \zeta_1} & \cdots & \frac{\partial f_n}{\partial \zeta_m} \end{pmatrix}.$$

The equations (3.1) are called the *normal equations* of the Least Squares problem.

4 Solving the normal equations

If $(\boldsymbol{\beta}_l, \boldsymbol{\zeta}_l, \boldsymbol{\lambda}_l)$ is an approximate solution to the normal equations, a refined solution $(\boldsymbol{\beta}_{l+1}, \boldsymbol{\zeta}_{l+1}, \boldsymbol{\lambda}_{l+1})$ can be found by the iteration

$$\begin{pmatrix} \boldsymbol{\beta}_{l+1} \\ \boldsymbol{\zeta}_{l+1} \\ \boldsymbol{\lambda}_{l+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_l \\ \boldsymbol{\zeta}_l \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \Delta\boldsymbol{\beta}_l \\ \Delta\boldsymbol{\zeta}_l \\ \boldsymbol{\lambda}_{l+1} \end{pmatrix} \quad , \quad l = 1, \ldots, \infty.$$

The step $(\Delta\boldsymbol{\beta}_l, \Delta\boldsymbol{\zeta}_l, \boldsymbol{\lambda}_{l+1})$ is given by

$$\mathbf{D}(\boldsymbol{\beta}_l, \boldsymbol{\zeta}_l) \begin{pmatrix} \Delta\boldsymbol{\beta}_l \\ \Delta\boldsymbol{\zeta}_l \\ \boldsymbol{\lambda}_{l+1} \end{pmatrix} = \begin{pmatrix} \mathbf{0}^{(k,1)} \\ \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\zeta}_l) \\ -\mathbf{f}(\boldsymbol{\beta}_l, \boldsymbol{\zeta}_l) \end{pmatrix}, \tag{4.1}$$

where

$$\mathbf{D}(\boldsymbol{\beta}_l, \boldsymbol{\zeta}_l) = \begin{pmatrix} \mathbf{0}^{(k,k)} & \mathbf{0}^{(k,m)} & \nabla_{\boldsymbol{\beta}}\mathbf{f}(\boldsymbol{\beta}_l, \boldsymbol{\zeta}_l)^T \\ \mathbf{0}^{(m,k)} & \boldsymbol{\Sigma}^{-1} & \nabla_{\boldsymbol{\zeta}}\mathbf{f}(\boldsymbol{\beta}_l, \boldsymbol{\zeta}_l)^T \\ \nabla_{\boldsymbol{\beta}}\mathbf{f}(\boldsymbol{\beta}_l, \boldsymbol{\zeta}_l) & \nabla_{\boldsymbol{\zeta}}\mathbf{f}(\boldsymbol{\beta}_l, \boldsymbol{\zeta}_l) & \mathbf{0}^{(n,n)} \end{pmatrix} \tag{4.2}$$

is a symmetric matrix. This iteration procedure is similar to Newton iteration except that the second order partial derivatives of the functions f_i have been neglected as it is practice to do in non-linear Least Squares estimation [4].

In order to reduce the effects of numerical rounding errors, it is recommended to calculate the step $(\Delta\boldsymbol{\beta}_l, \Delta\boldsymbol{\zeta}_l, \boldsymbol{\lambda}_{l+1})$ by solving the linear equations (4.1) by Gauss-Jordan elimination with full pivoting [4]. This algorithm also provides the inverse matrix $\mathbf{D}(\boldsymbol{\beta}_l, \boldsymbol{\zeta}_l)^{-1}$ which is needed at the final stage for estimating the covariance matrix of the solution as shown in Section 5.

If proper starting values $(\boldsymbol{\beta}_1, \boldsymbol{\zeta}_1)$ are selected, the iteration is expected to converge towards the solution $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})$

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\zeta}} \\ \boldsymbol{\lambda} \end{pmatrix} = \lim_{l \to \infty} \begin{pmatrix} \boldsymbol{\beta}_l \\ \boldsymbol{\zeta}_l \\ \boldsymbol{\lambda}_l \end{pmatrix}.$$

Since the solutions $\hat{\boldsymbol{\zeta}}$ are expected to be close to the estimates $\mathbf{z}$ of $\boldsymbol{\zeta}$ available a priori, the estimates $\mathbf{z}$ are obviously the proper starting values $\boldsymbol{\zeta}_1$ to be selected for the iteration. The selection of proper starting values $\boldsymbol{\beta}_1$ is more difficult in general. If, however, $\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta})$ are linear functions in the variables $\boldsymbol{\beta}$, the iteration process will converge after a few iterations, independent of the choice of $\boldsymbol{\beta}_1$.

Most differentiable functions $\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta})$ can be handled by the described method. In order to get reliable standard uncertainties, it is required, however, that the function can be approximated by a first order Taylor expansion, i.e.

$$\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta}) \cong \mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}) + \nabla_{\boldsymbol{\beta}}\mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \nabla_{\boldsymbol{\zeta}}\mathbf{f}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})(\boldsymbol{\zeta} - \hat{\boldsymbol{\zeta}})$$

when the values $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ are varied around the solution $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\zeta}}$ on a scale comparable to the standard uncertainties of the solution. If this vaguely formulated criterion is met, the function $\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta})$ is said to be *linearizable*. Note that almost any differentiable function

will be linearizable if the standard uncertainties are sufficiently small. On the other hand, if the uncertainties are sufficiently high, all non-linear functions will no longer be linearizable. The requirement that $\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta})$ is linearizable is considered to be the only major limitation of the method of Least Squares!

It should be mentioned that the minimization using Lagrange multipliers will fail in case the gradients $\nabla_\beta f_i$ and $\nabla_\zeta f_i$ of one of the constraint functions f_i are both equal to zero at the point of the solution $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})$. This gives some restrictions on how a constraint may be formulated. A function f_i defining a constraint may, for example, be replaced by the square of that function, f_i^2. But since $f_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}) = 0$, the gradient of f_i^2 will be zero at the point of the solution $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})$ although the gradient of f_i is not.

5 Properties of the solution

Since the solution $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}, \boldsymbol{\lambda})$ depends on the estimates $\mathbf{z}$, which are considered as the realization of the multivariate random variable $\mathbf{Z}$, the solution $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}, \boldsymbol{\lambda})$ can also be regarded as a multivariate random variable. If the functions $f_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})$ are linearizable, the estimates $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})$ are linear in $\mathbf{Z}$

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\zeta}} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\zeta} \\ \mathbf{0} \end{pmatrix} + \mathbf{D}(\boldsymbol{\beta}, \boldsymbol{\zeta})^{-1} \begin{pmatrix} \mathbf{0}^{(k,1)} \\ \boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \boldsymbol{\zeta}) \\ \mathbf{0}^{(n,1)} \end{pmatrix}. \tag{5.1}$$

In that case, the expectation of the solution is

$$E \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\zeta}} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\zeta} \\ \mathbf{0} \end{pmatrix}$$

which means that $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})$ are central estimators of the values $(\boldsymbol{\beta}, \boldsymbol{\zeta})$. Under the same assumption, the covariances of the solution are given by the symmetric matrix $\mathbf{D}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})^{-1}$ provided by the Gauss-Jordan elimination algorithm[2]

$$\begin{pmatrix} u(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^T) & u(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}^T) & ()^{(k,n)} \\ u(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}}^T) & u(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\zeta}}^T) & ()^{(m,n)} \\ ()^{(n,k)} & ()^{(n,m)} & -u(\boldsymbol{\lambda}, \boldsymbol{\lambda}^T) \end{pmatrix} = \mathbf{D}(\boldsymbol{\beta}, \boldsymbol{\zeta})^{-1} \cong \mathbf{D}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})^{-1}. \tag{5.2}$$

This relation can be derived as follows. Partition the symmetric matrix $\mathbf{D}^{-1}$ into nine sub-matrices similar to the left hand side of (5.2) or similar to the partitioning of $\mathbf{D}$ according to the definition (4.2). Express the covariance matrix of the solution (5.1) in terms of the covariance matrix $\boldsymbol{\Sigma}$ of the random variable $\mathbf{Z}$ and the matrix $\mathbf{D}^{-1}$. Insert the partitioned $\mathbf{D}^{-1}$ into the resulting matrix double product and express the covariances of the solution in terms of $\boldsymbol{\Sigma}^{-1}$ and the sub-matrices of $\mathbf{D}^{-1}$. Reduce these expressions to the final result by multiple use of the nine relations between the sub-matrices of $\mathbf{D}^{-1}$ and $\mathbf{D}$ derived from the identity $\mathbf{D}\mathbf{D}^{-1} = \mathbf{I}$.

[2]The empty brackets in the left hand matrix indicates the parts of $\mathbf{D}^{-1}$ that do not contain information about covariances.

From equation (5.1) and (5.2) the covariance matrices between $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})$ and the estimates $\mathbf{z}$ are found to be

$$u(\mathbf{z}, \hat{\boldsymbol{\beta}}^T) = u(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\beta}}^T)$$
$$u(\mathbf{z}, \hat{\boldsymbol{\zeta}}^T) = u(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\zeta}}^T).$$

From the last of these two relations, a relation of particular interest is derived,

$$u(\mathbf{z} - \hat{\boldsymbol{\zeta}}, \mathbf{z}^T - \hat{\boldsymbol{\zeta}}^T) = u(\mathbf{z}, \mathbf{z}^T) - u(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\zeta}}^T).$$

For the diagonal elements, this relation reads

$$u^2(z_i - \hat{\zeta}_i) = u^2(z_i) - u^2(\hat{\zeta}_i) \quad , \quad i = 1, \ldots, m.$$

That is, the variance of the difference between the initial estimate z_i of ζ_i and the refined estimate $\hat{\zeta}_i$ is equal to the variance of z_i minus the variance of $\hat{\zeta}_i$. This relation is useful when testing if the difference $z_i - \zeta_i$ is significantly different from its zero expectation.

6 χ^2 test for consistency

When the estimates $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})$ have been found, the minimum χ^2 value

$$\chi^2(\hat{\boldsymbol{\zeta}}; \mathbf{z}) = (\mathbf{z} - \hat{\boldsymbol{\zeta}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \hat{\boldsymbol{\zeta}})$$

can be used to test if the measured values $\mathbf{z}$ are consistent with the measurement model (2.1) within the uncertainties defined by the covariance matrix $\boldsymbol{\Sigma}$. If the model is linearizable, the expectation of the random variable $\chi^2(\hat{\boldsymbol{\zeta}}; \mathbf{Z})$ is equal to the number m of measured quantities, minus the number $m + k$ of adjusted quantities, plus the number n of constraints, that is

$$E\left[\chi^2(\hat{\boldsymbol{\zeta}}; \mathbf{Z})\right] = m - (m + k) + n = n - k = \nu.$$

If, in addition, the random variables $\mathbf{Z}$ are assumed to follow a multivariate normal distribution with mean values $\boldsymbol{\zeta}$ and covariance matrix $\boldsymbol{\Sigma}$, the random variable $\chi^2(\hat{\boldsymbol{\zeta}}; \mathbf{Z})$ will follow a $\chi^2(\nu)$ distribution with $\nu = n - k$ degrees of freedom. In that case, the probability p of finding a χ^2 value larger than the value $\chi^2(\hat{\boldsymbol{\zeta}}; \mathbf{z})$ actually observed can be calculated from the $\chi^2(\nu)$ distribution

$$p = P\left\{\chi^2(\nu) > \chi^2(\hat{\boldsymbol{\zeta}}, \mathbf{z})\right\} = 1 - P\{\chi^2(\nu) \leq \chi^2(\hat{\boldsymbol{\zeta}}, \mathbf{z})\}.$$

If this probability p is smaller than a certain value α, the hypothesis that the measured values are consistent with the measurement model has to be rejected at a level of significance equal to α. As the result of measurements are normally quoted at a 95% level of confidence, an $\alpha = 5\%$ level of significance is a reasonable choice for the consistency test.

Although the assumption of a normal distribution of $\mathbf{Z}$ may not be fulfilled, it is suggested to carry out the test of consistency as described above anyway. This is justified by the fact that a value of $\chi^2(\hat{\boldsymbol{\zeta}}; \mathbf{z})$ significantly higher than the expectation ν indicates inconsistency no matter what the distribution of $\mathbf{Z}$ might be. The calculated probability

p simply describes *how* unlikely the observed χ^2 value is *if* a normal distribution is assigned to Z.

7 Normalized deviations

If the test described in the previous section leads to a rejection of the measurements, a tool for identifying the outlying measurements is desirable. A measured value z_i is defined as an outlier if the difference $z_i - \hat{\zeta}_i$ is significantly different from zero taking into account the standard uncertainty $u(z_i - \hat{\zeta}_i)$ of that difference. This leads to the introduction of the normalized deviation d_i defined by[3]

$$d_i = \frac{z_i - \hat{\zeta}_i}{u(z_i - \hat{\zeta}_i)} = \frac{z_i - \hat{\zeta}_i}{\sqrt{u^2(z_i) - u^2(\hat{\zeta}_i)}} \quad , \quad i = 1, \ldots, m.$$

The normalized deviation d_i has zero expectation and variance 1. A normalized deviation with $|d_i|$ larger than 2 or 3 is therefore rather unlikely no matter what the distribution of the random variable d_i might be.

If a multivariate normal distribution is assigned to $\mathbf{Z}$ and the model function $\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta})$ is linearizable, the normalized deviation d_i is normally distributed,

$$d_i \in N(0,1) \quad , \quad i = 1, \ldots, m.$$

In that case

$$P\{|d_i| > 2\} = 5\%,$$

and a measurement with $|d_i| > 2$ is therefore identified as an outlier at a 5% level of significance. It is suggested to use the criteria $|d_i| > 2$ to identify potential outliers even if the distribution assigned to $\mathbf{Z}$ is not normal.

8 Adjustment of a variance σ^2

If some values z_i have a common but unknown variance $u^2(z_i) = \sigma^2$, this variance can be estimated by adjusting σ^2 by an iterative procedure until the "observed" χ^2 value becomes equal to its expectation value ν

$$\chi^2(\hat{\boldsymbol{\zeta}}; \mathbf{z}) = (\mathbf{z} - \hat{\boldsymbol{\zeta}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \hat{\boldsymbol{\zeta}}) = \nu,$$

where the covariance matrix $\boldsymbol{\Sigma}$ is a function of the unknown variance σ^2. As the estimates $\hat{\boldsymbol{\zeta}}$ depends on the value assigned to σ^2, these estimates have to be updated together with the estimates $\hat{\boldsymbol{\zeta}}$ each time the value of σ^2 is changed during the iteration.

This way of estimating the unknown variance σ^2 leads to the well-known expression for the standard deviation in the case of a repeated measurement of a single quantity as shown in Section 13.

[3] If $u(z_i - \hat{\zeta}_i) = 0$, the difference $z_i - \hat{\zeta}_i$ will be zero as well and d_i may be set equal to zero. This situation occurs whenever there is no redundant information available regarding the value of the quantity ζ_i.

9 Example: Calibration of an analytical balance

An analytical balance with capacity Max=220 g, resolution d=0.1 mg, and built-in adjustment weight was calibrated by DFM in October 1999 during an inter-laboratory measurement comparison piloted by DFM. Two mass standards were used as reference standards. One of them was a traditional 200 g weight (named R200g) of known conventional mass value[4] m_R and density ρ_R. The other reference standard was a specially designed 200 g stack of weights consisting of four discs (named 100g, 50g, 25g and 25g*) machined from the same metal bar of known density ρ. The conventional mass values m_1, m_2, m_3, m_4 respectively of these four discs were not known a priori; only the conventional mass value $m_S = m_1 + m_2 + m_3 + m_4$ of the stack was known.

The calibration was performed by placing a weight combination at the weighing pan of the balance and by recording the corresponding average indication I in the display. A total of 18 weight combinations were used. Each weight combination was weighed 3 times from which the average indication was calculated. The calibration was repeated 4 times during a period of 10 days in which the inter-laboratory comparison took place. From these four calibrations, a grand average indication $I_i, i = 1, \ldots, 18$ was calculated for each of the 18 weight combinations specified in Table 1. The standard uncertainty of the grand average was estimated from the observed variation in indication over the four calibrations.

I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9
100g	100g	100g	100g	100g	100g	100g	100g	100g
50g	50g	50g	50g	50g	25g	25g*	25g	
25g	25g	25g	25g*		25g*			
25g*	25g*							
I_{10}	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}	I_{17}	I_{18}
50g	50g	50g	50g	25g	25g	25g*	R200g	R200g
25g	25g	25g*		25g*				
25g*								

TAB. 1. The weight combinations corresponding to the 18 balance indications I_i.

Due to the effect of air buoyancy, the balance indication depends not only on the mass of the weighed body, but also on the density of the body as well as the density of the air. When calibrated in air of known density a, the reference indication I_R of the balance corresponding to a load generated by a weight with conventional mass value m and density ρ is given by

$$I_R = m\left(1 - (a - a_0)\left(\frac{1}{\rho} - \frac{1}{\rho_0}\right)\right)$$

where a_0=1.2 kg/m^3 and ρ_0=8000 kg/m^3 are the reference densities of the air and the weight respectively to which the conventional value of mass refers. As a model for the

[4]The conventional mass value of a body is defined as the mass of a hypothetical weight of density 8000 kg/m^3 that balances the body when weighed in air of density 1.2 kg/m^3 and temperature 20 ^{0}C.

	m_S [g]	m_R [g]	ρ_R [kg/m^3]	ρ [kg/m^3]	a [kg/m^3]	I_1 [div]
z	199.988816	199.999043	7833.01	7965.76	1.1950	199.988617
$u(z)$	0.000010	0.000008	0.29	0.71	0.0035	0.000023
$\hat{\zeta}$	199.988814	199.999043	7833.01	7965.76	1.1946	199.988620
$u(\hat{\zeta})$	0.000010	0.000008	0.29	0.71	0.0035	0.000011
d	1.66	-1.66	1.66	-1.66	1.66	-0.16

	I_2 [div]	I_3 [div]	I_4 [div]	I_5 [div]	I_6 [div]	I_7 [div]
z	199.988608	174.992133	175.009992	150.013558	149.980675	125.002083
$u(z)$	0.000023	0.000023	0.000023	0.000023	0.000023	0.000023
$\hat{\zeta}$	199.988620	174.992149	175.010024	150.013558	149.980672	125.002087
$u(\hat{\zeta})$	0.000011	0.000012	0.000012	0.000013	0.000012	0.000014
d	-0.56	-0.77	-1.61	0.03	0.14	-0.20

	I_8 [div]	I_9 [div]	I_{10} [div]	I_{11} [div]	I_{12} [div]	I_{13} [div]
z	124.984217	100.005650	99.982925	74.986433	75.004325	50.007892
$u(z)$	0.000023	0.000023	0.000023	0.000023	0.000023	0.000023
$\hat{\zeta}$	124.984212	100.005632	99.982899	74.986450	75.004325	50.007881
$u(\hat{\zeta})$	0.000014	0.000013	0.000013	0.000014	0.000014	0.000012
d	0.25	0.93	1.38	-0.87	0.03	0.54

	I_{14} [div]	I_{15} [div]	I_{16} [div]	I_{17} [div]	I_{18} [div]
z	49.974992	24.978533	24.996417	199.998867	199.998875
$u(z)$	0.000023	0.000023	0.000023	0.000023	0.000023
$\hat{\zeta}$	49.974995	24.978557	24.996432	199.998851	199.998851
$u(\hat{\zeta})$	0.000013	0.000011	0.000011	0.000011	0.000011
d	-0.19	-1.17	-0.77	0.78	1.19

	f [g/div]	A [1/div]	m_1 [g]	m_2 [g]	m_3 [g]	m_4 [g]
$\hat{\beta}$	1.00000186	-4.4E-09	100.005774	50.007963	24.978601	24.996476
$u(\hat{\beta})$	0.00000019	1.0E-09	0.000011	0.000010	0.000010	0.000010

TAB. 2. Measured and estimated values and associated standard uncertainties.

calibration curve of the balance, a second order polynomial through zero is assumed

$$I_R = f\left(I + AI^2\right)$$

where f and A are unknown quantities to be determined from the calibration data.

In this example, there are $m = 23$ quantities for which prior information is available from the measurements performed:

$$\boldsymbol{\zeta} = (m_S, m_R, \rho_R, \rho, a, I_1, \ldots, I_{18})^T$$

whereas there are $k = 6$ quantities for which no prior information is available:

$$\boldsymbol{\beta} = (f, A, m_1, m_2, m_3, m_4)^T.$$

	f	A	m_1	m_2	m_3	m_4
f	1	-0.945	0.021	0.071	0.096	0.096
A	-0.945	1	0.124	-0.016	-0.094	-0.094
m_1	0.021	0.124	1	-0.194	-0.269	-0.268
m_2	0.071	-0.016	-0.194	1	-0.287	-0.287
m_3	0.096	-0.094	-0.269	-0.287	1	-0.287
m_4	0.096	-0.094	-0.268	-0.287	-0.287	1

TAB. 3. Correlation coefficients of the estimated $\hat{\beta}$ values.

Between these quantities, there are $n = 19$ constraints:

$$\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta}) = \begin{pmatrix} (m_1 + m_2 + m_3 + m_4)\left(1 - (a - a_0)\left(\frac{1}{\rho} - \frac{1}{\rho_0}\right)\right) - f\left(I_1 + AI_1^2\right) \\ \vdots \\ m_R\left(1 - (a - a_0)\left(\frac{1}{\rho_R} - \frac{1}{\rho_0}\right)\right) - f\left(I_{18} + AI_{18}^2\right) \\ m_S - (m_1 + m_3 + m_3 + m_4) \end{pmatrix} = \mathbf{0} \ .$$

The measured values $\mathbf{z}$ and associated standard uncertainties are given in Table 2 under the row headings z and $u(z)$. All measured values are assumed to be uncorrelated.

By solving the normal equations, the estimates $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{\beta}}$ and associated standard uncertainties given in Table 2 under the row headings $\hat{\zeta}$, $u(\hat{\zeta})$, $\hat{\beta}$ and $u(\hat{\beta})$ are obtained. Selected correlation coefficients derived from $\mathbf{D}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})^{-1}$ are given in Table 3. The observed minimum χ^2 value is $\chi(\hat{\boldsymbol{\zeta}}, \mathbf{z}) = 8.6$ which should be compared to the expectation value $\nu = n - k = 19 - 6 = 13$. Since $P\left\{\chi^2(13) > 8.6\right\} = 80.3\%$, it is concluded that the measured values are consistent with the specified constraints taking into account the measurement uncertainties. This conclusion is confirmed by the calculated normalized deviations given in Table 2 under the row heading d; all normalized deviations satisfy the criterion $|d| < 2$.

From the estimates of the quantities f and A and the associated covariance matrix, the error of indication E, defined as

$$E = I - I_R = I - f\left(I + AI^2\right),$$

and the associated standard uncertainty $u(E)$ can be calculated as a function of the indication I. The result is shown in Figure 1 as the full lines representing $E - u(E)$, E, and $E + u(E)$. The measured points $E_i, i = 1, \ldots, 18$ shown in the figure are the observed average balance indications I_i minus the corresponding reference values I_R. The error bars of the measured points indicate the standard uncertainties $u(E_i)$ that have been calculated taking into account the covariance between I_i and I_R.

10 Example: Evaluation of calibration history

A weight (named R1mg) of nominal mass 1 mg has been calibrated 39 times in the period 1992-2001. For calibration number i, the mass m_i of the weight at the time t_i and the associated standard uncertainties $u(m_i)$ and $u(t_i)$ are given. The calibration history of the weight is shown in Figure 2 as dots with error bars indicating the standard

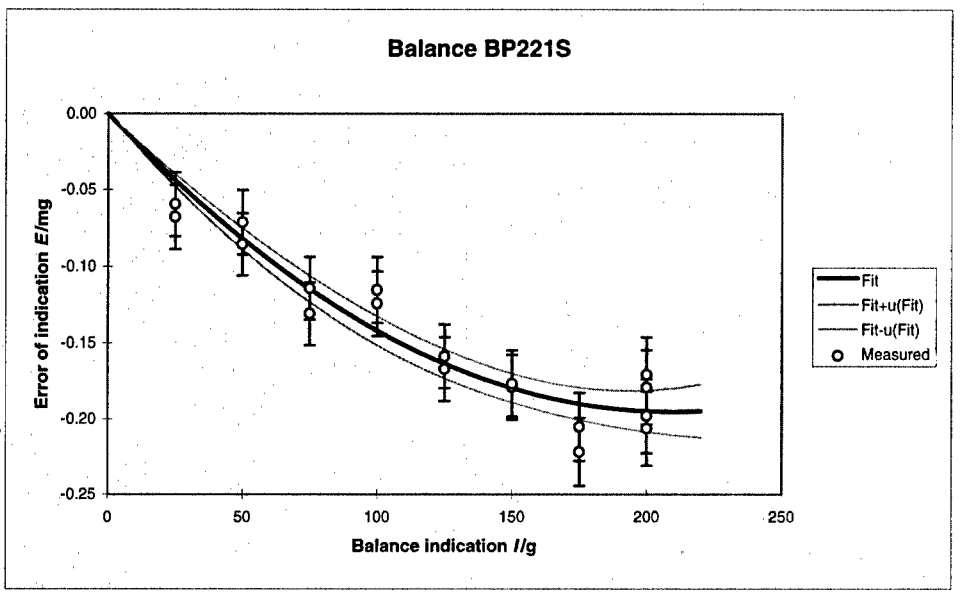

FIG. 1. Error of indication of the calibrated balance.

uncertainties; the scale mark 1992-01 on the time axis indicates the position of the date 1 January 1992 etc.

Due to wear and changes in the amount of dirt adsorbed to the surface, the mass of the weight is expected to change in time. A reasonable model of the change in mass as a function of time is a superposition of a deterministic linear drift and a random variation

$$m_i = a_1 + a_2 t_i + \delta m_i \quad , \quad i = 1, \ldots, 39,$$

where δm_i is a random variable with zero expectation and variance σ^2. The drift parameters a_1, a_2 and the associated covariance matrix as well as the variance σ^2 are unknown a priori and are to be estimated from the calibration history available. Once the estimates $\hat{a}_1$ and $\hat{a}_2$ have been found, it is possible to predict a value $\hat{m}$ of the mass of the weight as a function of time t

$$\hat{m} = \hat{a}_1 + \hat{a}_2 t + \hat{\delta m},$$

where $\hat{\delta m} = 0$ with standard uncertainty $u(\hat{\delta m}) = \sigma$. The standard uncertainty of the predicted mass value is given by

$$u^2(\hat{m}) = u^2(\hat{a}_1) + t^2 u^2(\hat{a}_2) + 2tu(\hat{a}_1, \hat{a}_2) + \sigma^2.$$

The measurement model used for evaluating the calibration history is

$$\boldsymbol{\zeta} = (m_1, \ldots, m_{39}, t_1, \ldots, t_{39}, \delta m_1, \ldots, \delta m_{39})^T \quad , \quad \boldsymbol{\beta} = (a_1, a_2)^T,$$

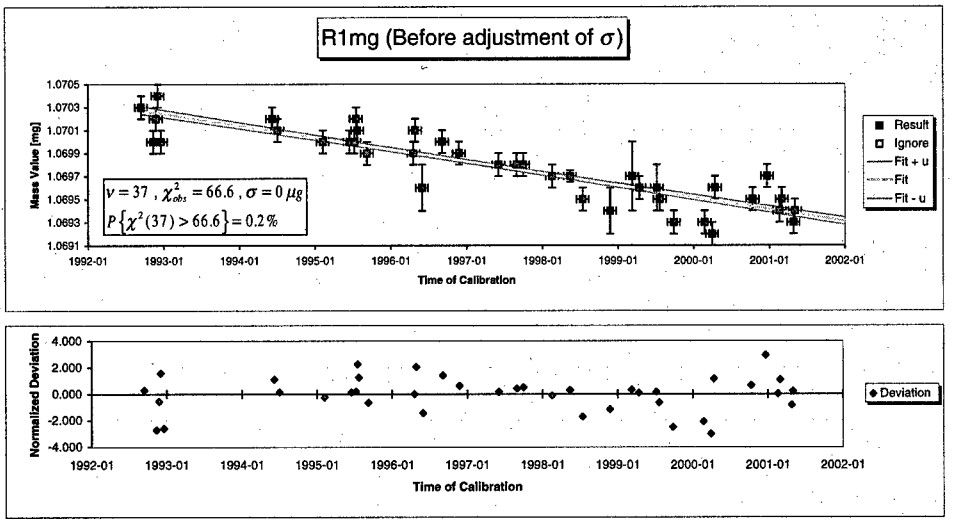

FIG. 2. Evaluation of the calibration history of a 1 mg weight assuming that $\sigma = 0$.

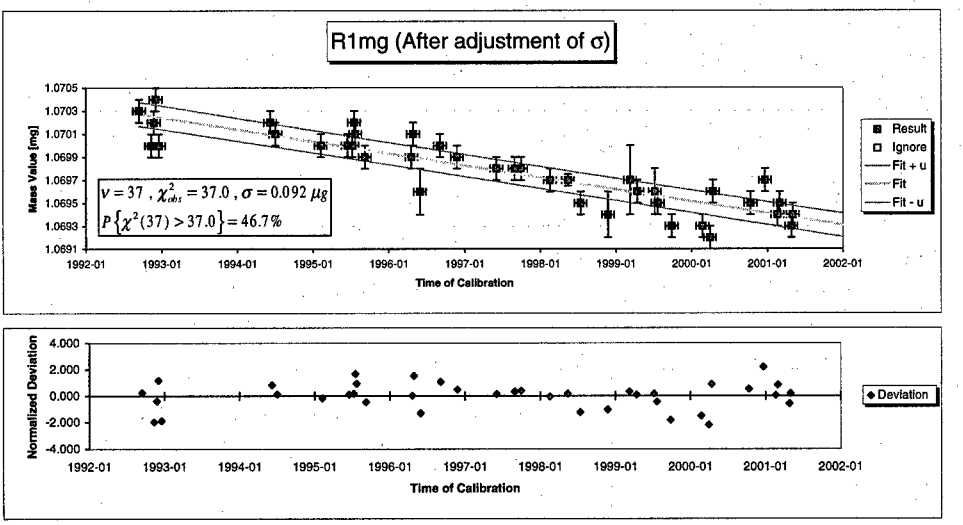

FIG. 3. Evaluation of the calibration history of the 1 mg weight with σ adjusted to 0.092 μg.

$$\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta}) = \begin{pmatrix} m_1 - (a_1 + a_2 t_1 + \delta m_1) \\ \vdots \\ m_{39} - (a_1 + a_2 t_{39} + \delta m_{39}) \end{pmatrix} = \mathbf{0}.$$

The measured values $\mathbf{z}$ are given by the calibration history, except for the values of

$\delta m_i, i = 1, \ldots, 39$ which are set equal to the expectation value zero. The associated covariance matrix $u(\mathbf{z}, \mathbf{z}^T) = \mathbf{\Sigma}$ is built up from the uncertainties $u(m_i)$ and $u(t_i)$ available from the calibration history and a negligible but finite[5] initial value of the unknown variance σ^2. Since the standard uncertainties $u(m_i)$ are of the order 0.1 μg, the value σ=1E-07 μg is considered negligible and is selected as a starting point.

By solving the normal equations, estimates $\hat{a}_1$ and $\hat{a}_2$ of the drift parameters and the associated covariance matrix are found after a few iterations. The predicted value $\hat{m}$ of the mass of the weight and the associated standard uncertainty $u(\hat{m})$ as a function of time are shown in Figure 2 as solid lines. The normalized deviations d associated with the mass values m_i are shown in Figure 2 as well[6]. The observed minimum chi-square value is $\chi^2 = 66.6$ which is large compared to the expectation value $\nu = 39 - 2 = 37$. Since $P\{\chi^2(37) > 66.6\} = 0.2\%$, the hypothesis $\sigma = 0$, or no random variation in the mass, is rejected at a 0.2% level of significance.

The value of σ is therefore increased as described in Section 8 until the calculated minimum χ^2 value becomes equal to its expectation value $\nu = 37$. In this way the standard uncertainty reflecting the random variation of the mass of the weight is found to be σ=0.092 μg. The result of the evaluation of the calibration history after adjustment of σ is shown in Figure 3. Note the significant increase in the standard uncertainty of the predicted value of the mass of the weight and the decrease in the absolute value of the normalized deviations d.

The calibration history can also be evaluated by an iterative technique based on linear regression [3]. The results obtained are identical to the results presented in this section.

11 Case I: Univariate output quantity, $Y = h(X_1, \ldots, X_N)$

In this section it is shown that the evaluation of measurements by the method of least squares is consistent with the generally accepted principles for evaluating measurement uncertainty as described in the GUM [1].

Using the nomenclature of the GUM, a univariate output quantity Y is assumed to be related to N input quantities $X_1, \ldots, X_N$ through a specified function h,

$$Y = h(X_1, \ldots, X_N).$$

The values assigned to the input and output quantities are denoted $x_1, \ldots, x_N$ and by y respectively. In the nomenclature of this paper, the measurement model is

$$\boldsymbol{\zeta} = (X_1, \ldots, X_N)^T \quad , \quad \boldsymbol{\beta} = (Y),$$

$$\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta}) = (Y - h(X_1, \ldots, X_N)) = 0.$$

The measured values are

$$\mathbf{z} = (x_1, \ldots, x_N)^T$$

[5] If the variance σ^2 is assumed to be exactly zero, the quantities δm_i have to be removed from the model. Otherwise the covariance matrix $\mathbf{\Sigma}$ will be singular.

[6] The absolute value of normalized deviations of t_i and δm_i is equal to the absolute value of the normalized deviation of m_i.

with the known covariance matrix

$$\mathbf{\Sigma} = u(\mathbf{z}, \mathbf{z}^T) = \begin{pmatrix} u^2(x_1) & \cdots & u(x_1, x_N) \\ \vdots & \ddots & \vdots \\ u(x_N, x_1) & \cdots & u^2(x_N) \end{pmatrix}.$$

The coefficient matrix $\mathbf{D}$ of the normal equations is

$$\mathbf{D}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}) = \begin{pmatrix} 0 & \mathbf{0}^{(1,N)} & 1 \\ \mathbf{0}^{(N,1)} & \mathbf{\Sigma}^{-1} & -\nabla_{\mathbf{x}} h(\mathbf{x})^T \\ 1 & -\nabla_{\mathbf{x}} h(\mathbf{x}) & 0 \end{pmatrix},$$

where

$$\nabla_{\mathbf{x}} h = \left(\frac{\partial h}{\partial X_1}, \ldots, \frac{\partial h}{\partial X_N} \right).$$

In the present case, the solution to the normal equations is found after one iteration,

$$y = \hat{\boldsymbol{\beta}} = h(x_1, \ldots, x_N) \quad , \quad \hat{\boldsymbol{\zeta}} = (x_1, \ldots, x_N)^T, \quad \lambda = 0.$$

The associated covariances are given by

$$\begin{pmatrix} u^2(y) & u(y, \hat{\boldsymbol{\zeta}}^T) & ()^{(1,1)} \\ u(\hat{\boldsymbol{\zeta}}, y) & u(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\zeta}}^T) & ()^{(N,1)} \\ ()^{(1,1)} & ()^{(1,N)} & -u^2(\lambda) \end{pmatrix} = \mathbf{D}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})^{-1}$$

$$= \begin{pmatrix} \nabla_{\mathbf{x}} h(\mathbf{x})\, \mathbf{\Sigma}\, \nabla_{\mathbf{x}} h(\mathbf{x})^T & \nabla_{\mathbf{x}} h(\mathbf{x})\, \mathbf{\Sigma} & 1 \\ \mathbf{\Sigma}\, \nabla_{\mathbf{x}} h(\mathbf{x})^T & \mathbf{\Sigma} & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

In other words,

$$u^2(y) = \nabla_{\mathbf{x}} h(\mathbf{x})\, \mathbf{\Sigma}\, \nabla_{\mathbf{x}} h(\mathbf{x})^T = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i u(x_i, x_j) c_j, \quad c_i \equiv \frac{\partial h}{\partial X_i}(x_i)$$

which is identical to the linear variance propagation formula given in the GUM.

12 Case II: Linear regression, $\mathbf{Y} = \mathbf{Xa}$

Linear regression is applied when there is a linear relationship $\mathbf{Y} = \mathbf{Xa}$ between some observed quantities $\mathbf{Y}$ and some unknown quantities $\mathbf{a}$. The *design matrix* $\mathbf{X}$ is made up of known elements that may be given as specified functions of one or several independent variables. In the notation of this paper, the measurement model for the linear regression problem is

$$\boldsymbol{\zeta} = \mathbf{Y} = (Y_1, \ldots, Y_n)^T \quad , \quad \boldsymbol{\beta} = \mathbf{a} = (a_1, \ldots, a_k)^T,$$

$$\mathbf{f}(\boldsymbol{\zeta}, \boldsymbol{\beta}) = \mathbf{Y} - \mathbf{Xa} = \mathbf{0},$$

where $\mathbf{X}^{(n,k)}$ is the known design matrix. The measured values are

$$\mathbf{z} = \mathbf{y} = (y_1, \ldots, y_n)^T$$

with known covariance matrix

$$\mathbf{\Sigma} = u(\mathbf{z}, \mathbf{z}^T) = \begin{pmatrix} u^2(y_1) & \cdots & u(y_1, y_n) \\ \vdots & \ddots & \vdots \\ u(y_n, y_1) & \cdots & u^2(y_n) \end{pmatrix}.$$

The coefficient matrix $\mathbf{D}$ of the normal equations is

$$\mathbf{D}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}) = \begin{pmatrix} \mathbf{0}^{(k,k)} & \mathbf{0}^{(k,n)} & -\mathbf{X}^T \\ \mathbf{0}^{(n,k)} & \mathbf{\Sigma}^{-1} & \mathbf{I}^{(n,n)} \\ -\mathbf{X} & \mathbf{I}^{(n,n)} & \mathbf{0}^{(n,n)} \end{pmatrix}.$$

Again, the solution to the normal equations is found after one iteration,

$$\hat{\mathbf{a}} = \hat{\boldsymbol{\beta}} = \mathbf{C}\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{y} \quad , \quad \hat{\mathbf{Y}} = \hat{\boldsymbol{\zeta}} = \mathbf{X}\hat{\mathbf{a}} \quad , \quad \boldsymbol{\lambda} = -\mathbf{\Sigma}^{-1}(\hat{\mathbf{Y}} - \mathbf{y})$$

where $\mathbf{C} \equiv (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$. The associated covariances are given by

$$\begin{pmatrix} u(\hat{\mathbf{a}}, \hat{\mathbf{a}}^T) & u(\hat{\mathbf{a}}, \hat{\mathbf{Y}}^T) & ()^{(k,n)} \\ u(\hat{\mathbf{Y}}, \hat{\mathbf{a}}^T) & u(\hat{\mathbf{Y}}, \hat{\mathbf{Y}}^T) & ()^{(n,n)} \\ ()^{(n,k)} & ()^{(n,n)} & -u(\boldsymbol{\lambda}, \boldsymbol{\lambda}^T) \end{pmatrix} = \mathbf{D}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})^{-1}$$

$$= \begin{pmatrix} \mathbf{C} & \mathbf{C}\mathbf{X}^T & -\mathbf{C}\mathbf{X}^T\mathbf{\Sigma}^{-1} \\ \mathbf{X}\mathbf{C} & \mathbf{X}\mathbf{C}\mathbf{X}^T & \mathbf{I} - \mathbf{X}\mathbf{C}\mathbf{X}^T\mathbf{\Sigma}^{-1} \\ -\mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{C} & \mathbf{I} - \mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{C}\mathbf{X}^T & \mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{C}\mathbf{X}^T\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1} \end{pmatrix},$$

that is,

$$\hat{\mathbf{a}} = \mathbf{C}\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{y} \quad , \quad u(\hat{\mathbf{a}}, \hat{\mathbf{a}}^T) = \mathbf{C} = (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$$

as is known from the theory of linear regression.

13 Case III: Repeated observations of a single quantity

Assume that a quantity X is measured n times with the same uncertainty σ. Such a measurement can be modelled by n quantities $X_1, \ldots, X_n$ having a common value μ

$$\boldsymbol{\zeta} = \mathbf{X} = (X_1, \ldots, X_n)^T \quad , \quad \boldsymbol{\beta} = (\mu),$$

$$\mathbf{f}(\boldsymbol{\beta}, \boldsymbol{\zeta}) = \begin{pmatrix} X_1 - \mu \\ \vdots \\ X_n - \mu \end{pmatrix} = \mathbf{0}.$$

The measured values are

$$\mathbf{z} = \mathbf{x} = (x_1, \ldots, x_n)^T,$$

and under the assumption that the measurement results are mutually independent, the associated covariance matrix is given by

$$\mathbf{\Sigma} = u(\mathbf{z}, \mathbf{z}^T) = \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix}.$$

The coefficient matrix $\mathbf{D}$ of the normal equations is

$$\mathbf{D}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}) = \begin{pmatrix} \mathbf{0}^{(1,1)} & \mathbf{0}^{(1,n)} & -\mathbf{1}^{(1,n)} \\ \mathbf{0}^{(n,1)} & \sigma^{-2}\mathbf{I}^{(n,n)} & \mathbf{I}^{(n,n)} \\ -\mathbf{1}^{(n,1)} & \mathbf{I}^{(n,n)} & \mathbf{0}^{(n,n)} \end{pmatrix},$$

where $\mathbf{1}$ denotes a matrix with all elements equal to 1. The solution of the normal equations is found after one iteration,

$$\hat{\mu} = \hat{\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \quad \hat{\mathbf{X}} = \hat{\boldsymbol{\zeta}} = \hat{\mu}\mathbf{1}^{(n,1)} \quad , \quad \boldsymbol{\lambda} = -\sigma^{-2}(\hat{\mathbf{X}} - \mathbf{x}).$$

The associated covariances are given by

$$\begin{pmatrix} u^2(\hat{\mu}) & u(\hat{\mu}, \hat{\mathbf{X}}^T) & ()^{(1,n)} \\ u(\hat{\mathbf{X}}, \hat{\mu}) & u(\hat{\mathbf{X}}, \hat{\mathbf{X}}^T) & ()^{(n,n)} \\ ()^{(n,1)} & ()^{(n,n)} & -u(\boldsymbol{\lambda}, \boldsymbol{\lambda}^T) \end{pmatrix} = \mathbf{D}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}})^{-1}$$

$$= \begin{pmatrix} \sigma^2 n^{-1} & \sigma^2 n^{-1}\mathbf{1}^{(1,n)} & n^{-1}\mathbf{1}^{(1,n)} \\ \sigma^2 n^{-1}\mathbf{1}^{(n,1)} & \sigma^2 n^{-1}\mathbf{1}^{(n,n)} & \mathbf{I}^{(n,n)} - n^{-1}\mathbf{1}^{(n,n)} \\ n^{-1}\mathbf{1}^{(n,1)} & \mathbf{I}^{(n,n)} - n^{-1}\mathbf{1}^{(n,n)} & \sigma^{-2}(n^{-1}\mathbf{1}^{(n,n)} - \mathbf{I}^{(n,n)}) \end{pmatrix}.$$

As expected,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \quad u^2(\hat{\mu}) = \sigma^2/n.$$

If σ^2 is not known a priori, it can be estimated by solving the equation

$$\chi^2(\hat{\boldsymbol{\zeta}}; \mathbf{z}) = \sum_{i=1}^{n} \frac{(x_i - \hat{\mu})^2}{\sigma^2} = n - 1,$$

i.e.,

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

which is the well known expression for the experimental standard deviation s.

14 Conclusion

A general technique for evaluation of measurements by the method of Least Squares has been presented. The applicability of the method has been demonstrated by two examples. It has been shown that the method is fully compatible with the generally accepted principles for evaluation of measurement uncertainty laid down in the GUM and that ordinary linear regression is just a special case of the method.

The **input** to the method consists of

- An estimate of the value of each measured quantity, including any relevant influence quantity.
- The covariance matrix of these estimates formed by the standard uncertainties of the estimates and the correlation coefficients between the estimates.

- A measurement model describing all the known relations between the measured quantities and some additional quantities (if needed) for which no prior information is available.

The **output** of the method consists of

- An adjusted estimate of the value of each measured quantity and an estimate of each additional quantity introduced in the measurement model.
- The covariance matrix of all these estimates from which the standard uncertainties and correlation coefficients can be calculated.
- A chi-square value which is a measure of the degree of consistency between the measurement model, the input estimates, and the covariances of the input quantities.

The adjusted estimate of the value of a measured quantity differs from the input estimate only if the measurement model imposes additional information regarding the value of that particular quantity. In that case the standard uncertainty of the adjusted estimate will be smaller than the standard uncertainty of the input estimate. For a good measurement, the difference between the adjusted estimate and the input estimate of a measured quantity should not be large compared to the standard uncertainty of that difference. It has therefore been suggested that the ratio d of the difference to its standard uncertainty is calculated and assessed against a selected criterion, e.g. $|d| < 2$. By plotting the d values of the adjusted estimates it is possible to assess whether a too high chi-square value is caused by a few poor input estimates or is due to a poor model.

Bibliography

1. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, *Guide to the expression of uncertainty in measurement*, ISO, 1995.
2. L. Nielsen, Least-squares estimation using Lagrange multipliers, *Metrologia* **35** (1998), 115–118. Erratum, *Metrologia* **37** (2000), 183.
3. L. Nielsen, Evaluation of the calibration history of a measurement standard, DFM report DFM-01-R25, 2001, 1–6.
4. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipies in C*, 2nd ed., Cambridge, Cambridge University Press, 1992, 36–40 and 681–688.
5. T. L. Saaty and J. Bram, *Nonlinear Mathematics*, Dover Publications, New York, 1981, 93–95.
6. K. Weise and W. Wöger, *Uncertainty and Measurement Data Evaluation*, Wiley-VCH, 1999, 183–224 [in German].